Realizing Organizational Collaboration through Semantic Mediation

Sri Gopalan, Sandeep Maripuri, Brad Medairy
Booz Allen Hamilton
8283 Greensboro Drive
McLean, VA 22102
703-902-5000
gopalan_sri@bah.com, maripuri_sandeep@bah.com, medairy_brad@bah.com

Abstract—Realizing[1,2] *organizational collaboration requires a greater level of information sharing between knowledge agents – both the people within an organization and the information systems that support them. Achieving this level of information transparency relies on fundamental improvements in today's systems and data mediation architectures. This paper describes how Semantic Web technologies can be leveraged within the context of Service Oriented Architectures to support dynamic, meaningful exchange of information both within and across organization boundaries.*

1. INTRODUCTION

Data interoperability, the "many to many exchanges of data that are sometimes predefined and sometimes unanticipated" [1], is a fundamental cornerstone of the Intelligence Community's (IC) drive towards information transparency [30]. However, today's enterprise environment faces many hurdles to achieve this level of information transparency: incongruous data representations, disparate and co-located data sources, stove-piped information conduits, and a general inability to understand what the data means or how it may be used. This information impedance creates a hostile environment for achieving information sharing within and across organizational boundaries, one of the IC's primary goals.

Currently, most of the work on this subject has focused on establishing a methodology for platform-neutral messaging and physical application connectivity via Web Services deployed within a Service Oriented Architecture (SOA). However, the development of a complementary data integration solution, which transforms the raw messages into meaningful, actionable information, lags significantly behind.

"Traditional" attempts to solve this data impedance problem, specifically mapping-based and shared schema-based approaches, suffer from significant limitations, saddling organizations with brittle, inflexible, and hard-wired solutions or with potentially inconsistent data representations with no facility to validate whether the enclosed data values actually express the intended meaning of each data element. Further analysis reveals that the key missing aspect of current data integration approaches is a mechanism to explicitly describe what the exchanged data means, and how it is intended to be used.

By exposing this information - the semantics of a data source – the data may be described beyond mere structure and syntax, enabling the proper consumption of the real-world concepts that the data source encodes. In support of a standardized semantic modeling language, the W3C Semantic Web technologies, like the Resource Description Framework (RDF) and Web Ontology Language (OWL), offer a codified, computing model to express the meaning encoded in disparate data formats [2].

This paper describes a novel approach towards achieving data interoperability within a SOA through the use of the Semantic Web technologies. Specifically, by examining the meaning of data elements, and using OWL ontologies to bridge disparate data element labels, aggregation schemas, and data usage, this semantic mapping technique enables data to be unambiguously expressed, complete with specific business rules regarding use, and rationalized with an overall knowledge model that bridges the different concepts hidden within the various data representations.

Once semantically mapped, by leveraging inferencing technology this approach can then dynamically merge, classify, and recast data in arbitrary formats meaningful to the end consumer in a more scalable, loosely-coupled manner than previously available. Ultimately, this approach enjoys the theoretical benefits of existing data integration solutions, without experiencing the same prohibitive implementation drawbacks.

Semantic interoperability enables effective information sharing within and across community boundaries, allowing organizations to bridge the gaps inherent in data integration exercises. This high level of information interoperability promotes the ability to dynamically discover, access, and consume information, enabling the IC effectively collaborate in mission-critical timeframes.

2. ORGANIZATIONAL COLLABORATION AND SOA

The IC is making strides towards becoming a "'smart[er]' government [that will] integrate all sources of information [from within and across organizational boundaries] to see the enemy as a whole" [3]. From an enterprise computing standpoint, this change has resulted in

[1] 0-7803-9546-8/06/$20.00© 2006 IEEE
[2] IEEEAC paper #1302, Version 3, Updated October 19, 2005

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **21 MAY 2008**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Realizing Organizational Collaboration through Semantic Mediation**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Booz Allen Hamilton 8283 Greensboro Drive McLean, VA 22102**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S)
		11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES **AFCEA-GMU C4I Center Symposium "Critical Issues In C4I" 20-21 May 2008, George Mason University, Fairfax, Virginia Campus, The original document contains color images.**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**29**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

a fundamental shift away from application-centric computing, where only a privileged set of users access data from isolated applications that perform a limited set of tasks, to a more situation-centric model, where an knowledge consumer interacts with a loosely-coupled system that provides dynamic, context-sensitive capabilities based on the operational needs of the problem at hand [4].

Essential to this vision of improved organizational collaboration is a greater ability to enable knowledge producers and knowledge consumers – both the people within an organization as well as the software systems that support them - to better coordinate the sharing of information.

From an enterprise architecture perspective, the physical infrastructure to provide the underpinnings of this data sharing platform, in part, lies in the adoption of a Service-Oriented Architecture (SOA), implemented using Web Services and other XML-driven initiatives [5].

Within a SOA, services are visible to the network at large by providing physical interfaces over enterprise assets. These services are platform neutral, and are described with application-agnostic service descriptions which can be published to metadata registries. Thus, network-enabled users are provided an open, standards-based means to discover relevant services and to invoke them either individually or within the framework of a larger composite process that leverages several services across the enterprise. SOA provides a foundational layer for an information-centric enterprise that satisfies new and changing business needs by enabling the dynamic sharing and aggregating of information across organizational boundaries via individual service-enabled enterprise assets [5].

However, SOA, by itself, is merely an abstract architecture specification; World Wide Web Consortium (W3C) [28] and the Organization for the Advancement of Structured Information Standards (OASIS) [29] endorsed Web Service standards (WS-*), built over XML and Web technologies, represent the latest attempt to realize the full capabilities of SOA. These standards provide many of the essential physical infrastructure components required of a SOA platform:

(1) ***Message Encoding:*** SOAP is a standardized specification for encoding message payloads between services.

(2) ***Service Interface Description:*** Web Service Description Language (WSDL) describes a Web service's capabilities as collections of communication endpoints capable of exchanging messages.

(3) ***Service Metadata Registry:*** Universal Description, Discovery, and Integration (UDDI) is a registry for services to expose their interface descriptions for discovery on the network.

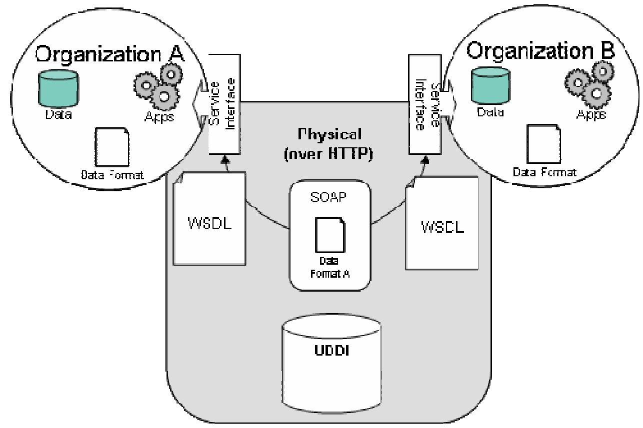


Figure 1 - SOA Realized with Web Services and XML technologies [33]

While the WS-* initiatives facilitate the discovery and delivery of data, they do little to actually improve upon the exchange of information. By definition, data are merely physical values, while information is the contextual interpretation of data that gives it meaning [6]. This vital aspect of the overall SOA solution requires the ability to properly interpret of raw message data into comprehendible meaning.

3. Data Interoperability

Data interoperability aims to fill this gap. In contrast to data integration, forcibly fitting multiple systems together to form a singular unit, interoperability requires that systems be able to work together without expending special effort to ensure they understand each other directly [6]. Integration typically requires brittle, hard-coded data transformations between each system participating in the data exchange scenario – they must be aware of one another, and modifications to any one system has a cascading affect to all other systems that may interact with it. Interoperability, on the other hand, attempts a more loosely coupled approach by promoting operation isolation of each system [27]. In this context, systems only need to be aware of their data representations – the solution-at-large should handle mediating representational differences between system interactions.

Accomplishing this task requires that SOA infrastructure be able to make sense of the various data representations in use, and be able to sensibly negotiate transformations and resolve any potential data conflicts.

Data representation interpretation is typically managed through the evaluation of metadata – data that is used to describe the meaning and usage of other data [7]. Metadata can be generally classified as follows:

Table 1 – Types of Metadata

Metadata Type	Description	Example
Syntactic	Describes the syntactic markup of data	Datatype, field length
Structural	Describes the aggregation of multiple individual data elements into larger, composite record units	Physical schema descriptions (PersonRecord; PersonName)
Semantic	Describes the codified relationships between data elements, including any rules or constraints on those relationships	Person was-born on PersonDOB, and was-born only once

With these different types of metadata, each describing data at a different level of abstraction, data interoperability schemes have a vehicle to holistically describe data assets. Solutions then leverage each of these metadata types in varying degrees to achieve their specific vision of data interoperability.

While metadata facilitates data description, however, it does not by itself completely obviate the three main classes of risk associated with data heterogeneity: schematic conflicts, semantic conflicts, and intensional conflicts.

Schematic conflicts, the most typical problems encountered in data interoperability, arise when trying to combine multiple sources of data that may model or structurally organize data differently [8].

Table 2 – Schematic Data Conflicts

Conflict Type	Description
Data Type	Different primitive system types to represent the same piece of data (xsd:datetime vs. TIMESTAMP)
Label	Similar concepts labeled differently (CUSTOMER vs. PURCHASER)
Aggregation	Same set of related information aggregated and related differently (same number of aggregations and relations, different verb phrases for relationships and thus different aggregation values)
Generalization	Different levels of abstractions for same type of data

Semantic conflicts root from “the fact that data present in different systems may be subjected to different interpretations.” [8] Generally speaking, these types of issues are prevalent when common schemas are used. In these cases, the data may be schematically conformant, but the misinterpretation of schematic organization leads to value-based disjoints.

Table 3 – Semantic Data Conflicts

Conflict Type	Description
Naming	Same concept expressed with different values (BAH vs. Booz Allen vs. Booz Allen Hamilton). Sometimes referred to as “Value Normalization”
Scaling	Different units of measurement to express same concept (Grade of “A” expressed as 4.0 vs. 5.0)

Intensional conflicts refer to fundamental differences between the informational content supplied by the data producer and the expectation of the data consumer [8].

Table 4 – Intensional Data Conflicts

Conflict Type	Description
Domain	Differing interpretation regarding actual domain being modeled (model provides stock performance profile summary; one implementation includes S&P 500, second includes entire Dow Jones index)
Integrity Constraint	Differing integrity constraints asserted in multiple systems (similar concept exists uniquely in one system – allowing it to be used as a key element, but may be repeated in another system)

Effective solutions carefully address these issues by balancing responsibility across the different entities that play a role in enabling interoperability: the communities that act as interpreters, describing what the data means, individual organizations who act as guardians, physically managing the data, and the software systems that act as messengers, facilitating the transfer of the data between different parties.

4. TRADITIONAL APPROACHES

Traditional data interoperability approaches have, for the most part, focused on the standardization or manipulation of syntactic and structural metadata to achieve their mediation goals. Domain-specific, standardized message formats and XML standards-based transformation, two of today’s predominant solutions, demonstrate the great strengths and apparent weaknesses representative of almost all data interoperability solutions to date [9].

4.1. Domain-Specific Standardized Message Formats

One of the oldest approaches to addressing the data mediation challenge focuses on community developed, domain-specific standardized message formats, like RosettaNet [10] or Intelligence Community Markup Language (ICML) [11]. These standards, driven by the needs of its community members, hope to facilitate speed, efficiency, and reliability of message transfers and enable greater communication and collaboration amongst trading partners. In theory, a common and consistent data representation promises some significant benefits:

(1) A controlled vocabulary describing the semantic meaning of data elements for all community members

(2) The expected syntax and structure of data elements is concisely expressed without ambiguity and may be externally validated

(3) A standardized format facilitates automated processing of messages with minimal human interaction

In the context of addressing the primary types of data conflicts, standardized formats rely heavily on the uptake of community standards and the diligence of individual organizations to conform to these specifications. In this scenario, information systems only take on the minor role of data type validation.

Table 5 – Shared-Schema Responsibility of Data Conflicts

Entity	Data Conflicts Addressed	Description
Community Standards	Labeling, Aggregation, Generalization, Naming, Scaling	Standards rigidly defined how data should look and what formats its content should follow
Individual Organizations	Confounding, Domain, Integrity Constraint	Organizations are the only entity that can ensure that the information sent is correct
Information Systems	Data Type	Systems can only validate syntax and structure, not meaning

In practice, however, this methodology breaks down in several areas. As with any community-driven effort, building consensus across a wide range of stakeholders is generally a slow and politically driven process dominated by the larger players. This may endanger the data needs of smaller partners, leading to certain compromises in data classification and possible data infidelity.

Secondly, the widespread acceptance of any standard depends on garnering critical mass within the domain of use. Since adopting standardized-schemas is an all-or-nothing approach, if the requisite level of participation has not been reached, then potential partners may be hesitant to join without a clear idea of the standard's ability to gain widespread industry acceptance.

Finally, the machine-to-machine interoperability benefits do not scale when two domain-related, but structurally different, specifications are incorporated together. From an interoperability standpoint, the primary value of shared schema approaches derives from the standardized syntax and structure of the specification, not the underlying meaning the data expresses. Even though semantic metadata plays a vital role in aligning the community around a common understanding of those concepts represented by structural elements, a shared vocabulary is difficult to leverage for true interoperability. For instance, the Global Justice XML Data Model (GJXDM) initiative attempts to merge several justice-related schemas into one, all-encompassing, super-schema [12]. The experience of integrators trying to implement GJXDM has been that while it may be possible for humans to infer equivalence between disparate data elements based on their semantic descriptions, it is virtually impossible for today's information systems to perform the same logical operation based structure alone [13]. As a result, integrators must deal with each physical difference independently as they arise, thus lessening the benefit of automation promised by the approach.

4.2. XML standards-based transformation

As XML processing standards, like XSL/T, XPath, and XQuery, have matured, many organizations have managed data transformations via XML transformation. These routines can be coded to handle the representational differences between disparate data formats.

Armed with XML transformation, individual organizations are not obligated to conform to any particular data specification a priori and can "map" their data to other representations at a later date. This offers some significant advantages over the shared schema approach:

(1) Allows organizations to remain internally-focused and rapidly develop formats that are specifically engineered to solve their particular data needs

(2) Standards-based mechanism to allow organizations to recast their data into virtually any other data representation for interoperability

(3) Organizations do not have to completely abandon proprietary, legacy specifications to participate in an information-sharing network

This said, however, by not focusing on a set standard, the XML transformation methodology shifts most of the burden of resolving data conflicts from the community to the information systems supporting interoperability.

Table 5 – XML Transform Responsibility of Data Conflicts

Entity	Data Conflicts Addressed	Description
Community Standards	N/A	Transformations may be advised by standards, but are not subject to them
Individual Organizations	Confounding, Domain, Integrity Constraint	Organizations must deal with differences in domain definitions and relationships and find the "right" place to put data within a document
Information Systems	Data Type, Labeling, Aggregation, Generalization, Naming, Scaling	Systems can only validate syntax and structure, not meaning

While this approach offers a well-accepted solution to bridge different data representation together, there is no computable means of asserting that the output of the transform has not altered the meaning of the original data in any way. If the integrator responsible for creating the mapping does not completely understand the meaning expressed in the target specification, then there is a high possibility of incorrect associations and incongruent data.

In addition, this mechanism breaks down within the dynamics of a service-oriented architecture. Given the multitude of available services and data formats, which could conceivably be arranged in any number of

permutations, this static, point-to-point approach towards coded data mappings requires a transformation routine to be created for every possible permutation.

With N number of data formats, this requires N^2 - N mappings. While certainly manageable with four or five different formats, this model becomes unmanageable even when the number of formats approaches ten. Furthermore, because these mappings are brittle, point-to-point integrations, they duplicate invasive data interpretation rules in each mapping. In other words, any small change in one data format could break any mappings within which that format participates, requiring up to 2 x N - 2 mapping modifications. While well-adopted, this approach moves towards an integration mindset, and abandons the loose coupling tenets of interoperability.

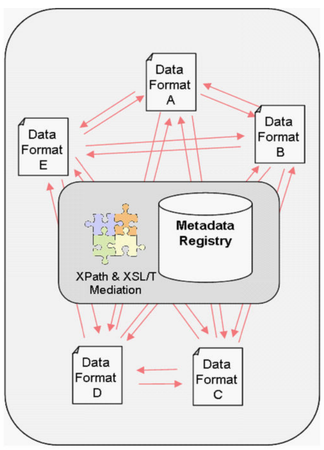


Figure 2 – Point to Point Mediation [33]

4.3. Assessment of Traditional Approaches

Shared schema's strengths lie mainly with knowledge consumers, who depend on a consistent data format with pre-defined meanings for its elements. Conversely, XML transformations primarily benefits knowledge producers who are able more specifically define specifications that may extend and tailor some of the concepts expressed in community-endorsed formats, but represent them differently. However, both of these approaches do not suit the needs of an information-centric environment where the needs of a potentially large and diverse group of both the knowledge consumer and knowledge producer must be met.

5. Computable Semantics

Semantics is the shared meaning of data within the context of a business domain—the business concepts comprising a business domain and their explicit inter-relationships. The use of semantics enables access to information within a context. Traditionally, approaches to data interoperability have not explicitly captured semantics. The meaning of data is implied within the mappings and generated code; but it is not externalized or computable. What the data actually means, or how it is used, is still largely the burden of subject matter experts, data stewards, and rules encoded into programs. The "smarts" is in software, transformation scripts, or in people's heads, and not in metadata or reusable models.

An improved data interoperability methodology requires smart, reusable models. By encoding the meaning and usage of data in interpretable metadata descriptions, point-to-point mappings can be avoided. Software will be able to interpret message formats based on the business concepts contained within, enabling dynamic aggregation and transformation of data. Thus, this proposed semantic mediation approach intends to improve upon existing data interoperability techniques by performing data mediation at the semantic, rather than rather than syntactic or structural, level.

There are two important precursors to being able to fully implement semantic mediation:

(1) A computable-semantic model that describe information contents in an unambiguous, machine-interpretable manner

(2) The development of knowledge models describing the domains relevant to COI interest areas and operation

These two vital pieces, a computable-semantic model and the development of domain ontologies, are problems that are actively being addressed by various standards bodies and COIs respectively.

5.1. Computable Semantic Models and the Semantic Web

Semantic knowledge models capture the real-world "facts" regarding a business domain in a computable manner. Similar in function to Entity-Relationship or Object models that are commonly used in software engineering, semantic knowledge models explicitly capture the specific nature of relationships between entities or business concepts. Employing a knowledge model enables enterprises "to assert the domain of interest, and the relationships between the concepts that comprise the domain" [33]:

Table 6 – Knowledge model components [33]

Component	Description	Examples
Concept	Abstract business entities that may be realized by one or more actual things	'Terrorist', 'Event', 'Victim'
Relationship	The nature of connectedness between abstract business entities	'ParticipatesIn', 'OccursAt'
Constraint	Conditions required to satisfy the existence of a relationship between abstract business entities	Cardinality, Optionality, Nullability
Rule	Logical rules regarding concepts, relationships, and constraints	If A & B, Then C

As a specific type of knowledge model, an ontology takes the form of a graph structure where the nodes represent the business concepts within a business domain and the arcs represent the business relationships between those concepts (Table 6) [7]. Typically, most ontology languages provide three fundamental types of relationships to aid in the description of a business domain: equivalence, subsumption (inheritance), and disjointness.

(1) *Equivalence* allows foreign or differently named concepts to be asserted to be the same thing.

(2) *Subsumption* allows the specification of concept specialization—a sub-concept inherits the basic meaning and properties of a super-concept, but additionally participates in more relationships thereby specifying its meaning in a narrower context.

(3) *Disjointedness* allows the specification that two concepts are entirely incompatible, and that no realizations of those concepts could ever be classified as meaning the same thing.

In essence, ontologies allow for the development of a well-defined domain model that explicitly defines the concepts and relationships comprising that domain.

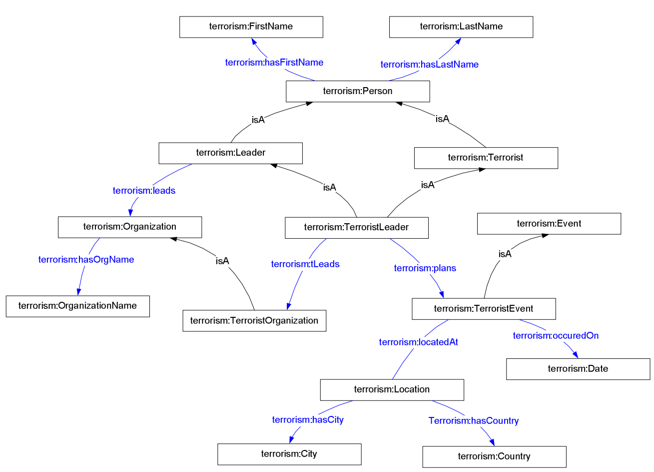


Figure 3 – Sample terrorism ontology [33]

Ontologies, because of their inherent graph structures, offer great flexibility and power from a computational perspective that empowers machines to interpret and reasoning against the models. For example, relationships between business concepts can be autonomously traversed by a computer to deduct unstated correlations between entities (Figure 3) allowing latent knowledge to be logically discovered. Furthermore, an ontology may include logical axioms that, when enforced, enable complex inferences and conclusions to be drawn against instance data values that previously might have gone unseen by a human operator.

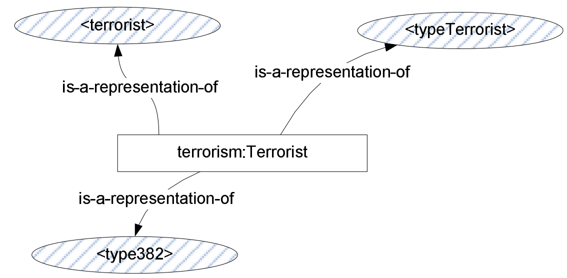


Figure 4 – Varying XML tags representing same concept [33]

The aforementioned built-in relationship types, including equivalence, subsumption, and disjointedness, as well as any custom logical relationship can not only be used to describe relations within a particular ontology, but may also be used to logically bridge concepts from one ontology to those in a different ontology. By asserting relationships between foreign concepts, data value relationships can also be inferred.

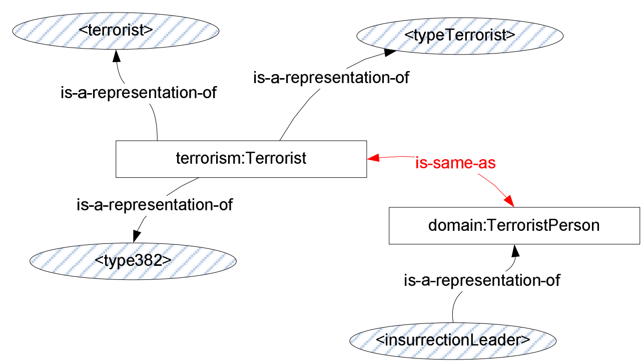


Figure 5 – Bridging Ontologies [33]

Furthermore, many ontology languages allow transitivity of the relationships between business concepts; the built-in relationship types are transitive by definition and any custom relationship may be explicitly declared as transitive. This allows undeclared relationships to be deduced across concepts that may not be directly linked. By logically connecting two ontologies, the pre-existing linkages between any other ontology to which either belongs may be inferred. This propagates, creating a large-scale network effect, ultimately decreasing the number of bridges that must be manually created between foreign ontological concepts.

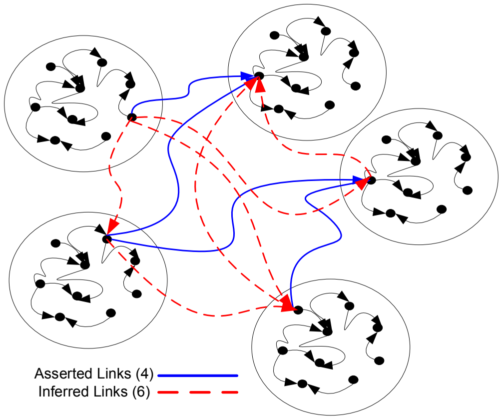


Figure 6 – Network effect of bridging ontologies [33]

Computable semantic models, a subject of research in the artificial intelligence community for over 20 years, manifest themselves in a variety of representation languages such as KIF, FLogic, and OCML. While many of these logic-based dialects are variants of first-order predicate calculus, where "reasoning amounts to verifying logical consequence" [34], many also support higher-order logics where the increased level of expressivity sometimes allows for the construction of statements that are neither complete, guaranteeing that all conclusions are computable, nor decidable, ensuring that all conclusions may be computed in finite time.

In support of a standardized document format for information capture and exchange, the W3C Semantic Web Activity has recommended three standardized document formats [14].

Table 7 – W3C Semantic Web specifications

Specification	Description
Resource Description Framework (RDF) [15]	A data model language for representing the relationships between resources ("actual things")
RDF Schema Language (RDFS) [16]	RDF-encoded language for representing the basic relationships between classes of resources ("types of actual things")
Web Ontology Language (OWL) [2]	RDF-encoded language, building over RDF Schema, for describing ontologies, including more expressive relationships, constraints, and rules

OWL provides three increasingly expressive sublanguages designed for use by specific communities of implementers and users: [2]

(1) **OWL Lite** supports those users primarily needing a classification hierarchy and simple constraints. For example, while it supports cardinality constraints, it only permits cardinality values of 0 or 1. It should be simpler to provide tool support for OWL Lite than its more expressive relatives, and OWL Lite provides a quick migration path for thesauri and other taxonomies. Owl Lite also has a lower formal complexity than OWL DL. [2]

(2) **OWL DL** supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL DL includes all OWL language constructs, but they can be used only under certain restrictions (for example, while a class may be a subclass of many classes, a class cannot be an instance of another class). OWL DL is so named due to its correspondence with description logics, a field of research that has studied the logics that form the formal foundation of OWL. [2]

(3) **OWL Full** is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. For example, in OWL Full a class can be treated simultaneously as a collection of individuals and as an individual in its own right. OWL Full allows an ontology to augment the meaning of the pre-defined (RDF or OWL) vocabulary. It is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full. [2]

Built over widely adopted XML and Web standards, these Semantic Web specifications facilitate integration with existing technologies in addition to providing standardized languages to encode formal semantic markup. It is important to note, however, that not all variants of OWL are applicable to the data mediation use cases.

Data modeling within a data mediation context requires logic formalisms that must be both computable and decidable; a scenario where a conclusion cannot logically be drawn is unacceptable. Thus, the less expressive logic dialects like OWL-Lite and OWL-DL, which are guaranteed to be conclusive, are appropriate while the more expressive OWL Full, which cannot ensure that assertions are complete or decidable, is conversely not acceptable.

But, even while confined to a less expressive form, semantic modeling and ontological reasoning can have a profound effect on data mediation. By overlaying a semantic definition to descriptions of data and services in machine-understandable formats, organizations are encouraged to continue to use data formats tailored to their needs while seamlessly allowing that same data to be computably interpreted and leveraged by the community at large.

5.2. Development of Domain Ontologies

To enable the description of physical data elements in a conceptual form, the ontologies that define the concepts and relationship within that domain must first be in place. Indirectly supporting this requirement, many COIs have been encouraged to "develop an ontology that best reflect the community understanding of their shared data." [1]

This drive to “enable data to be understandable” [1] coupled the W3C’s formal acceptance of OWL as its standard ontology description language, has reinvigorated the growth of domain ontology development within the government. As a result, many complementary organizations have formed to facilitate and support these ontology building activities.

The Semantic Interoperability Community of Practice (SICoP), sponsored by the Chief Information Officers Council (CIOC) in partnership with the XML Working Group, chartered itself with the purpose of “achieving ‘semantic interoperability’ and ‘semantic data integration’ focused on the government sector.” [17]. The Ontology and Taxonomy Coordinating Working Group (ONTACWG), a special working group within SICOP, focuses on promoting “collaborat[ion] in the actual construction of useful knowledge representation systems” and “interoperability by identifying common concepts among knowledge classifications developed by different groups”.[18]. The National Center for Ontological Research (NCOR), recently founded by various academic, commercial, and government entities, aims to “advanc[e] ontological investigation within the United States.”[19]

As the education, acceptance, and development of ontologies expands, because of the network effect, the descriptive power of these ontological models may stretch beyond the business domains that they were originally intended for. As a consumer of this semantic knowledge, the reach of semantic mediation, by extension, will also cross organizational boundaries, thereby allowing COIs to further break down their interoperability barriers.

6. SEMANTIC MEDIATION – AN APPROACH

With a computable-semantic model in place along with the emerging development of COI-specific domain models, the beginnings of semantic mediation are in place. This semantic mediation strategy currently focuses its efforts to give greater structure to using computable-semantics in data interoperability.

There are several methodologies proposed in literature that advocate the use of semantics in the context of data integration. The SAINT project approaches this data mediation problem with a “mediator-wrapper” architecture that effectively translates local RDMS data sources data into OWL and uses a global mediator to perform semantic translations [20]. The MAFRA toolkit focuses on the “lift and normalization” of source data formats and ontologies, as well as providing a methodology for instance transformation [21]. The Artemis initiative, which most closely tracks our goals, provides semantic interoperability in the healthcare domain by wrapping existing applications as Web Services, normalizing legacy EDI and XML formats into OWL, and using OWL-QL to mediate the semantic differences [22]. While these various efforts within the research community exhibit some commonality with this semantic mediation solution, in total they do not espouse the same goals this approach hopes to achieve.

6.1. Physical – Conceptual Round-tripping

For any XML data representation, an XML Schema Definition (XSD) describes the low-level document structure and content details. Similarly, the concepts and relationships implicit with the data representation can also be formally codified using OWL.

The ability to conceptually normalize the implicit semantic information hidden within XSD into OWL and, conversely, de-normalize that same conceptual OWL representation back down to its physical XSD form is crucial to the semantic mediation algorithm. This process may be described as physical – conceptual round-tripping [20]. To perform this action, a mapping linking the two types of metadata must be created. This Concept Mapping explicitly describes three distinct aspects of the implicit data model expressed within a particular data representation: Concept Entities, Concept Attributes, and Entity Bridges.

A Concept Entity is a complex-typed XML element that represents a higher level domain concept, such as `TerroristLeader` or `TerroristEvent`.

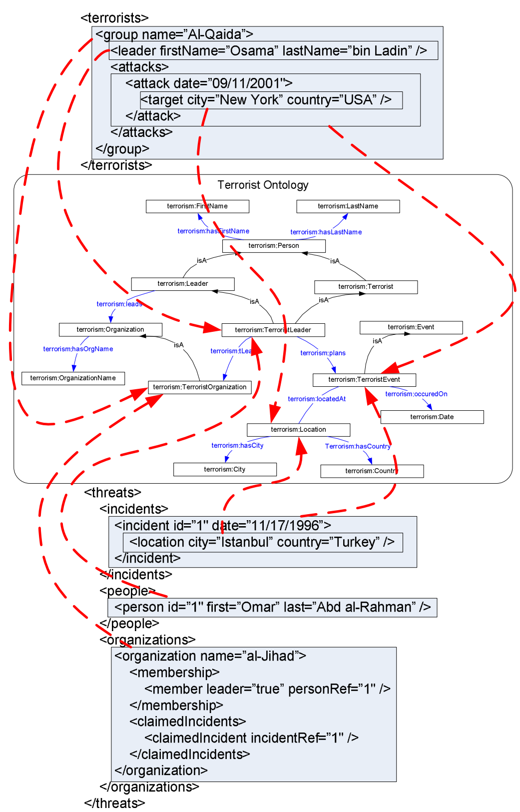


Figure 6 – Concept Entity mapping to Terrorist Ontology

A Concept Attribute is an XML attribute or a simple XML element that also represents a high-level domain concept, but has a physical value, such as `OrganizationName` or `City`. Concept Attributes are

explicitly linked as members of Concept Entities through higher level domain relationships such as `hasOrgName(Organization, Name)` or `hasCity(Location, City)`.

An Entity Bridge represents the higher level domain relationship between two roundtrip entities, such as `plans(TerroristLeader, TerroristEvent)`. An Entity Bridge also describes how to syntactically and structurally navigate from the XML element represented by one Concept Entity to the XML element that represents the Concept Entity it is related to. It is important to note that the domain ontology models relationships between concepts identically, regardless of whether the physical representation happens to be linked Concept Entities (Entity Bridge) or properties between a Concept Entity and a Concept Attribute. The Entity Bridge construct provides a reification over the former property to indicate that the physical serialization is two separate XML Concept Entities which are related.

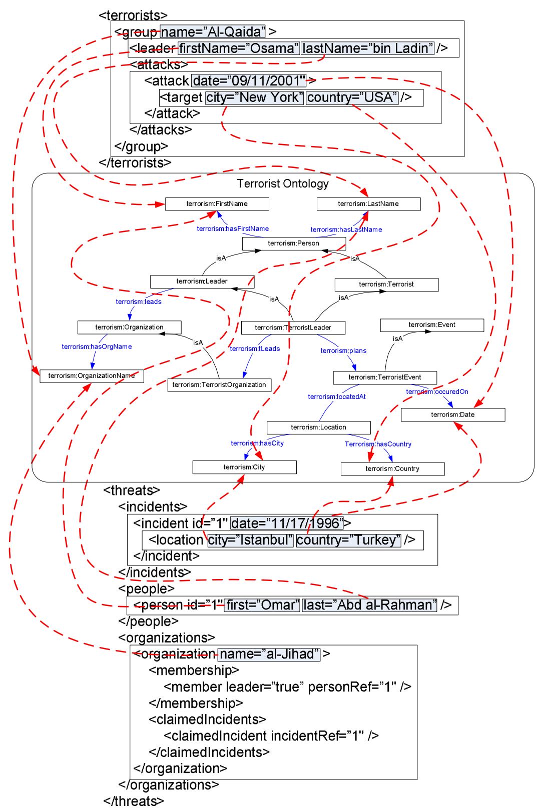


Figure 7 – Concept Attributes mapping to Terrorist Ontology

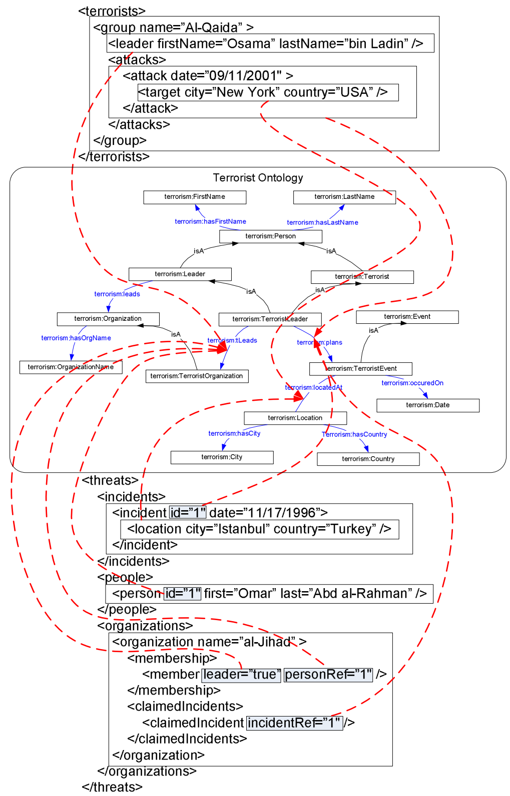


Figure 8 – Entity Bridges mapping to Terrorist Ontology

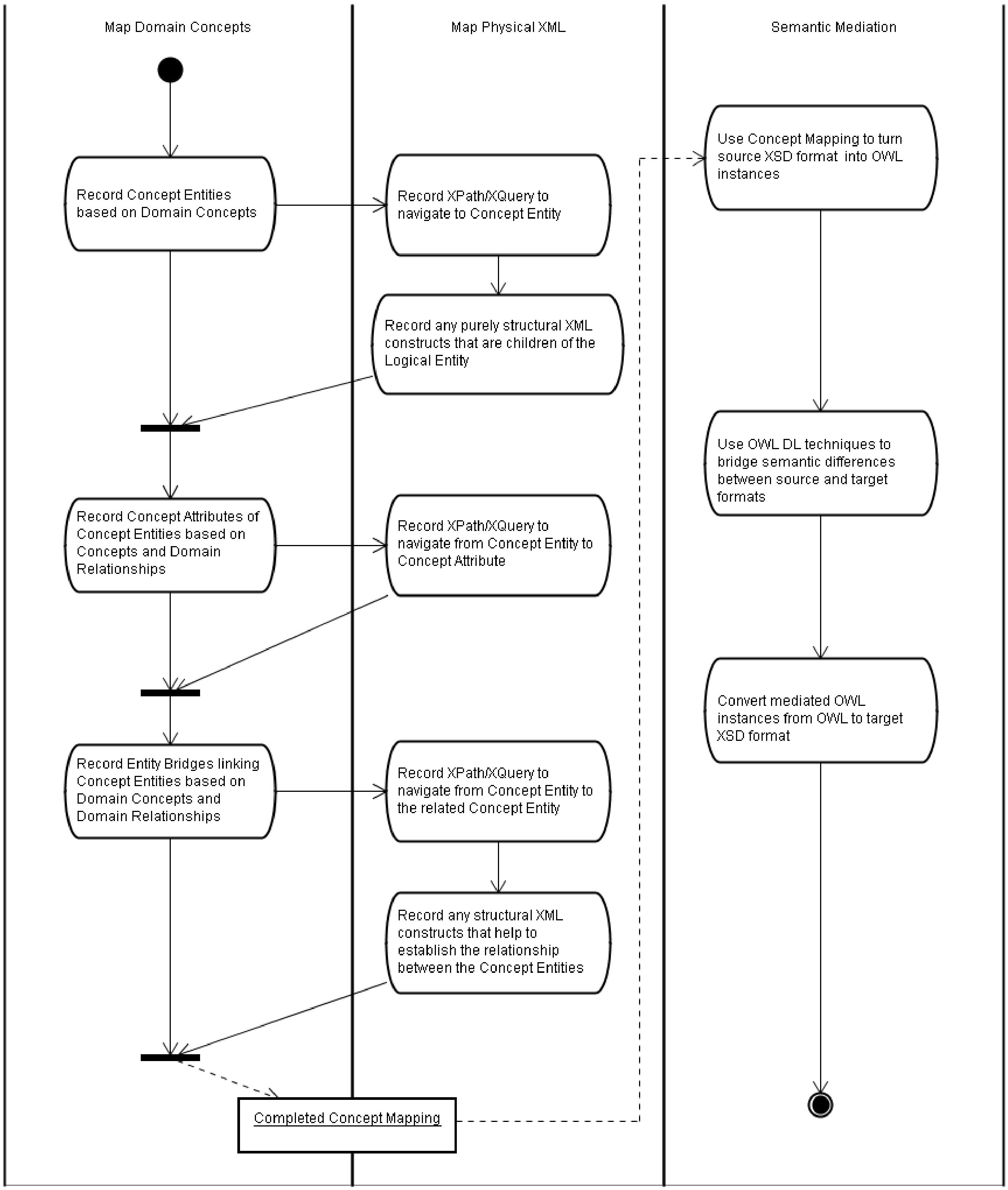


Figure 9 – Concept Mapping and Semantic Mediation

The general algorithm to extract this Context Mapping information from a XML data representation is as follows:

(1) Identify the Concept Entities within the XML document and document the OWL domain concept that they represent. Also document any child XML structures that have no logical value but are necessary to conform to the document schema.

(2) For each Concept Entity, identify the related Concept Attributes. Document the OWL domain concept each logical attribute represents as well as the OWL property that describes the relationship between the Concept Entity and the Concept Attribute. .

(3) For each Concept Entity, identify the Entity Bridges that relate them to other Concept Entities. Document the OWL property that describes the bridge relationship between the two Concept Entities. Also document any special XML structures that have no logical value, but are utilized to define the structural relationship between the two Context Entities.

(4) For each of the identified Concept Entities, Concept Attributes, and Entity Bridges, note the meta-information about the XML components that they represent, such as the name and whether it is an element, as well as the XPath or XQuery expression to reach one node from the one it is related to.

After having performed the Concept Mapping of both the source and target data representations, the semantic mediation may begin. To transform the physical XSD data representation of the source document into its corresponding conceptual form, the semantic mediation algorithm iterates over the Concept Entities, Concept Attributes, and Entity Bridges defined in the source Concept Mapping and creates corresponding OWL instances based on the related concepts and relationships defined within those mappings. In this process, the explicit data values within the source document are extracted via the pre-defined XPath and XQuery expressions and instantiated as an `owl:DataTypeProperty` [2] instance (named "`hasValue`") against the related OWL instance.

Once in a conceptual form, the semantic mediation algorithm walks the concept graph defined by the target Concept Mapping. For each concept and relationship encountered, the algorithm uses OWL DL reasoning and some advanced matching techniques to find the associated source OWL instance that satisfies the target mapping class. This includes mechanisms to determine class and property compatibility through equivalence and subsumption checking. Once the source OWL instances can be rationalized in the context of the target Concept Mapping, the de-normalization process into the target XSD data representation may begin.

This de-normalization process involves iterating over the Concept Entities, Concept Attributes, and Entity Bridges defined in the target Concept Mapping, creating the appropriate XML elements and XML attributes, and then merge those individual XML nodes together into a valid target XSD format.

Table 8 – OWL Instance Results

```
<rdf:RDF ... >
  <terrorist:TerroristLeader rdf:ID="#leader0">
    <terrorist:tLeads rdf:about="#org0" />
    <terrorist:plans rdf:about="#event0" />
    <terrorist:hasFirstName rdf:about="#first0" />
    <terrorist:hasLastName rdf:about="#last0" />
  </terrorist: TerroristLeader>
  <terrorist:TerroristOrganization rdf:ID="#org0">
    <terrorist:hasOrgName rdf:about="#name0" />
  </terrorist:TerroristOrganization>
  <terrorist:TerroristEvent rdf:ID="#event0">
    <terrorist:locatedAt rdf:about="#loc0" />
    <terrorist:occuredOn rdf:about="#date0" />
  </terrorist:TerroristEvent>
  <terrorist:Location rdf:ID="#loc0">
    <terrorist:hasCity rdf:about="#city0" />
    <terrorist:hasCountry rdf:about="#country0" />
  </terrorist:Location>
  <terrorist:OrganizationName rdf:ID="#name0">
    <mapping:hasValue rdf:datatype="&xsd;string">
         Al-Quida
    </mapping:hasValue>
  </terrorist:OrganizationName>
  <terrorist:Date rdf:ID="#date0">
    <mapping:hasValue rdf:datatype="&xsd;string">
         09/11/2001
    </mapping:hasValue>
  </terrorist:Date>
  <terrorist:City rdf:ID="#city0">
    <mapping:hasValue rdf:datatype="&xsd;string">
         New York
    </mapping:hasValue>
  </terrorist:City>
  <terrorist:Country rdf:ID="#country0">
    <mapping:hasValue rdf:datatype="&xsd;string">
         USA
    </mapping:hasValue>
  </terrorist:Country>
  <terrorist:FirstName rdf:ID="#first0">
    <mapping:hasValue rdf:datatype="&xsd;string">
         Osama
    </mapping:hasValue>
  </terrorist:FirstName>
  <terrorist:LastName rdf:ID="#last0">
    <mapping:hasValue rdf:datatype="&xsd;string">
        bin Ladin
    </mapping:hasValue>
  </terrorist:LastName>
</rdf:RDF>
```

7. Reflections and Future Directions

As we are in the early stages of our work, we have conducted a variety of experiments testing our methodology using existing XML and OWL tools. To develop sample OWL ontologies, SWOOP [23], a hypermedia based OWL ontology editor developed by the Mindswap group at the University of Maryland, College Park, was leveraged. Also, Pellet [24], the Mindswap group's Java-based OWL-DL reasoner, was used to validate many of basic assumptions regarding OWL reasoning. To integrate Pellet into our infrastructure, we leveraged Jena [25], a popular Java-based Semantic Web framework. Also, to facilitate the composition and decomposition of XML data elements

during our physical-conceptual round-tripping process, the Saxon XSL/T and XQuery processor API was used [26].

With these basic pieces, an early semantic mediation prototype has successfully achieved data interoperability between schematically simple source and target XSD representations with OWL-Lite conformant semantic descriptions.

From these preliminary results, this semantic mediation approach demonstrates many of the benefits of its traditional predecessor's approaches while filling in many of their gaps:

(1) Facilitates better runtime automation without advocating a particular syntax or structure

(2) Explicitly leverages well-defined domain concepts and relationships to perform mediation

(3) Leverages existing standards and specifications

(4) Extensible, scalable solution that is flexible to grow and change along with the altering data requirements and domain knowledge

In terms of mitigating the inherent data conflicts in data mediation, this approach takes the middle ground between the shared schema and XML transformation methodologies. While it relies on the community to define a flexible domain ontology, it shifts more of the processing focus away from the individual organizations and more towards the realm of information systems. This effectively, lessens the burden on the integrators within each organization and, potentially allowing them to work more efficiently.

Table 9 – XML Transform Responsibility of Data Conflicts

Entity	Data Conflicts Addressed	Description
Community Standards	Domain, Generalization, Naming, Scaling	Community developed OWL ontologies give offer flexible domain model to describe the important concepts and relationships
Individual Organizations	Confounding, Integrity Constraint	Organizations only have to focus on explicitly describing the semantic concepts implicit within their specific data formats
Information Systems	Data Type, Labeling, Aggregation,	The physical-conceptual round-tripping will allow organizational formats to be normalized in OWL and then computably reasoned against

This said, preliminary experimentation has focused on employing a single ontology to pivot between disparate physical representations of data. We recognize that in a true interoperability environment, multiple ontologies describing multiple domains would be employed; these different ontologies would have to be bridged to ensure cross-COI information exchanged. As part of this exercise, we have begun investing various mechanisms to provide ontology bridging. One mechanism would be to provide these bridges as separate OWL documents, utilizing OWL Lite and OWL DL properties, such as `owl:equivalentClass`, `owl:subClassOf`, `owl:disjointWith`, `owl:equivalentProperty`, `owl:subPropertyOf`, `owl:inverseOf`, complex class types, etc. [2], to assert relatedness between classes and properties in disparate ontologies. This approach has the benefit of pure conformance to OWL DL, and due to the use of in-built language constructs enables disparate ontologies and bridging documents to be classified and merged using an OWL DL reasoner. The net result: a large, virtual ontology, resident within the reasoner, which combines and relates all relevant ontology classes and properties.

We have also begun investigating various ontology mediation languages, including the SEKT Project's [31] Ontology Mediation Management language [32]. At this time, there does not appear to be an implementation of the abstract language, nor supporting software infrastructure to perform ontology merging.

We have also thus far assumed that all properties used in COI domain ontologies are instances of `owl:ObjectProperty` [2] – this approach does not currently support the use of `owl:DatatypeProperty` instances to describe attributes of domain ontology classes. This assumption is partially built on experiences building ontologies where content is entirely abstract: all elements within the ontology are either concepts or relationships. Further investigation will be required to better cope with the possibility of ontologists using `owl:DatatypeProperty` instances in their ontologies instead of `owl:ObjectProperty` instances, and potential difficulties in mediating between these two constructs during ontology bridging exercises.

8. CONCLUSIONS

Using ontologies allows local organizations and COIs to describe the meaning of their data explicitly, instead of encoding interpretation with respect to other COI message formats in mappings. Instead of brittle, static data mappings that are tied to the specific syntax of a particular data format, organizations can bridge the differences in their data at a conceptual level. Through this level of abstraction, changing the syntax of a particular field no longer invalidates other mappings. Further, due to the network effect implicitly available in OWL, mapping complexity grows linearly with the number of different data formats.

The research to date does not constitute a real-world, functioning system, but does highlight the promising benefits of the semantic mediation approach. Some open issues that are not discussed in this paper, like the exact algorithm for OWL instance comparison, syntactic data translations, service enablement, and performance issues due to the scaling up of ontologies are areas for future work.

References

[1] DoD Net-Centric Data Strategy, http://www.dod.mil/nii/org/cio/doc/Net-Centric-Data-Strategy-2003-05-092.pdf

[2] W3C OWL Web Ontology Language Reference, http://www.w3.org/TR/owl-ref/

[3] 9-11 Commission Report, http://www.gpoaccess.gov/911/

[4] Situational Software Blog, http://netmesh.info/blog/Situational%20Software/what-is-situational-software.html

[5] Margaret Arney, Brad Cohen, Brad Medairy, "Impact of Advanced Collaborative Architectures on Intelligence Analysis," 2004 IEEE Aerospace Conference Proceedings, March 2004

[6] Wikipedia, http://en.wikipedia.org/wiki/Data

[7] Jeffrey T. Pollack, Ralph Hodgson, "Adaptive Information", John Wiley & Sons, 2004 pp. 32-38, pp.127–134, pp. 148-149., pp. 184-187

[8] C.H. Goh, "Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Systems," Sloan School of Management, Massachusetts Institute of Technology, 16-22, January 1997

[9] Sandeep Maripuri, Thanikai Sokka, Brad Medairy, "Enabling Intelligence Analysis through Semantic Information Interoperability," 2005 IEEE Aerospace Conference Proceedings, March 2005

[10] RosettaNet, http://www.rosettanet.org/RosettaNet

[11] Intelligence Community Metadata Working Group, https://www.icmwg.org/

[12] Global Justice XML Data Model, http://it.ojp.gov/jxdm/

[13] Global Justice XML Data Model Developer's Workshop, http://justicexml.gtri.gatech.edu/developerworkshop/index.html

[14] World Wide Web Consortium Semantic Web Activity, http://www.w3.org/2001/sw

[15] W3C RDF Resource Description Framework (RDF): Concepts and Abstract Syntax, http://www.w3.org/TR/rdf-concepts/

[16] W3C RDF Vocabulary Description Language 1.0: RDF Schema, http://www.w3.org/TR/rdf-schema/

[17] Semantic Interoperability Community of Practice Wiki, http://colab.cim3.net/cgi-bin/wiki.pl?SICoP

[18] Ontology and Taxonomy Coordinating Working Group, http://colab.cim3.net/cgi-bin/wiki.pl?SICoP/OntologyTaxonomyCoordinatingWG

[19] National Center for Ontological Research, http://www.ncor.us/

[20] Patrick Lehti, Peter Fankhauser. "XML Data Integration with OWL: Experiences and Challenges," *saint*, p. 160, 2004 International Symposium on Applications and the Internet (SAINT'04), 2004

[21] Fodor, O., Dell'Erba, M., Ricci, F., Spada, A. and Werthner, H,; "Conceptual Normalisation of XML Data for Interoperability in Tourism, " Workshop on Knowledge Transformation for the Semantic Web (KTSW 2002) at ECAI'2002, pp. 69-76, 2002.

[22] Veli Bicer, Gokce B. Laleci, Asuman Dogac, Yildiray Kabac, "Artemis Message Exchange Framework: Semantic Interoperabilty of Exchanged Messages in the Healthcare Domain", http://www.srdc.metu.edu.tr/webpage/projects/artemis/publications/_MappingHL7v3-v2.pdf

[23] SWOOP, http://www.mindswap.org/2004/SWOOP/

[24] Pellet, http://www.mindswap.org/2003/pellet/

[25] Jena, http://jena.sourceforge.net/

[26] Saxon, http://saxon.sourceforge.net/

[27] Oussama Kassem Zein and Yvon Kermarrec, "An Approach for Describing/Discovering Services and For Adapting Them to the Needs of Users in Distributed Systems", American Association for Artificial Intelligence, 2004

[28] World Wide Web Consortium, http://www.w3.org/

[29] Organization for the Advancement of Structured Information Standards, http://www.oasis-open.org/

[30] Richard Lardner, "DOD Experiment Could Lead to Major Changes in Intel Operation," Inside the Army, 1-4, July 7, 2003

[31] Semantically Enabled Knowledge Technologies, http://sekt-project.org

[32] SEKT, "Ontology Mediation Management", http://sw.deri.org/~jos/deri-mediation/mediation-management.pdf

[33] Sri Gopalan, Sandeep Maripuri, Brad Medairy, "Realizing the Net-Centric Data Strategy with Semantic Web Technologies," Booz Allen Hamilton Journal of Applied Analytics in Consulting (BAHJAAC), Vol. 2, No. 1, January 2006, pp.12.

[34] F. Baader, D. Calvanese, D. L. McGuiness, D. Nardi, P. F. Patel-Schneider: "The Description Logic Handbook: Theory, Implementation, Applications", 2003.

BIOGRAPHY

Sri Gopalan is an Associate with Booz Allen Hamilton's Global IT Team, leading SOA design, governance, and interoperability efforts for Defense and Intelligence Community clients. He has over 7 years of professional experience developing and architecting enterprise-class applications and services for both the commercial and government sectors. He is currently serving as a development lead, researching and implementing various Semantic Web and SOA-based prototypes focused on promoting collaboration and information sharing for a classified project within the government. He holds a MS and BS in Electrical and Computer Engineering from Carnegie Mellon University.

Sandeep Maripuri is a Senior Associate with Booz Allen Hamilton's Global IT Team, leading Applied Research & Development efforts for Defense and Intelligence Community clients. His focus areas include applying advanced concepts (e.g. Semantic Web, Grid Computing) to operational needs and Net-Centric architectures. These efforts target methods for improving the efficiency and dynamic composability of large, distributed systems. He is currently overseeing the implementation of several Semantic Web and SOA-based prototypes focused on promoting collaboration, data interoperabiltiy, and information sharing for research-oriented clients. Prior to joining Booz Allen, Sandeep had provided consulting services in addition to working in the COTS marketplace, where he helped architect and build a semantics-based, dynamic data integration product. Sandeep holds a B.S. in Mechanical Engineering, minor Computer Science, from the University of Illinois at Urbana-Champaign.

Brad Medairy is a Principal with Booz Allen Hamilton's Global IT Team, headquartered in Mclean, VA. As a leader in the firm's Service Oriented Architecture (SOA) Solutions area, he leads a team focused on the strategy, design, and implementation of SOA and integration solutions. He has a proven track record in the application of emerging technologies (e.g. Semantic Web, Social Computing, Grid Computing, and Web Services) to address the business and missions needs of customers across all areas of Government. He holds an MS in Information Systems and Technology from Johns Hopkins University and a BS in Information Systems from University of Maryland, Baltimore County.

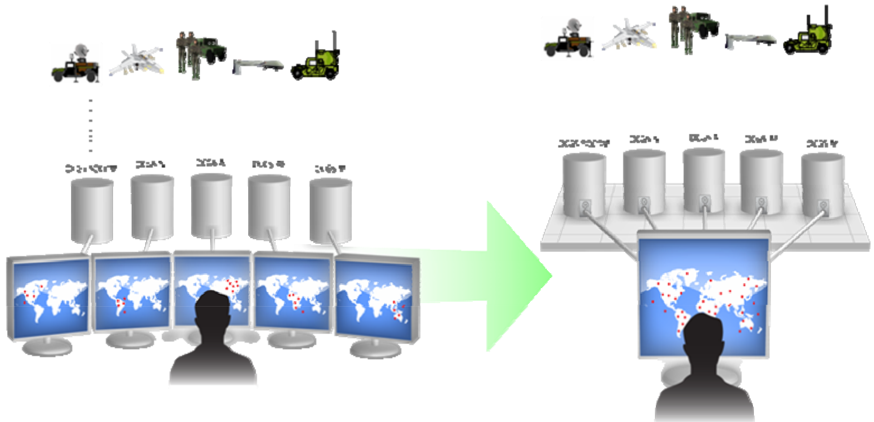
Realizing Organizational Collaboration through Semantic Mediation

An Approach for Dynamic Data Interoperability within the Intelligence, Surveillance, and Reconnaissance (ISR) Community

Agenda

- Data Mediation Challenges in the ISR COI
- Introduction to Computable Semantics
- Introduction to Semantic Mediation
- Applying Semantic Mediation to the ISR COI

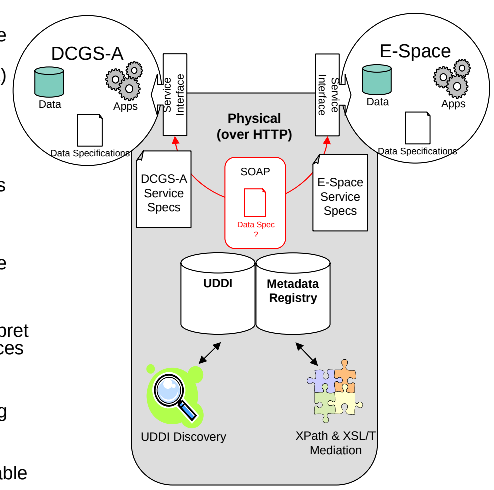
The objective state ISR operational view provides integrated battlespace awareness across multiple data assets regardless of sensor, platform, and organizational boundaries



- The realization of this vision requires the ability to exchange data in an interoperable fashion in addition to an improved capacity to understand information from a variety of sources

While the ISR Community has begun to embrace SOA to achieve organization-level information sharing, it has not completely addressed inter-organization interoperability

- Programs such as the Army's DCGS-A and the Intelligence Community's E-Space have embraced Service Oriented Architecture (SOA) concepts
  - Data Services have increased **internal** visibility and accessibility of data with Web Services and XML technologies
  - **Organization-level** data interoperability has been achieved through the use of **internal** data specifications
- Interoperability between DCGS-A and E-Space has not yet been completely achieved due to divergent data specifications
  - Analysts must be able to discover and interpret 3rd party specifications to find external sources of relevant data
  - 3rd party specifications must be mediated to resolve syntactic differences across differing specifications
  - Mediation infrastructure must scale to meet increased demands as the number of available service specifications increases



While Web Services and XML have addressed physical interoperability well, they are still challenged in providing scalable information interoperability solutions

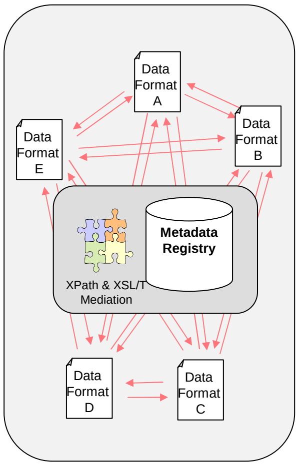
- The core Web Services and XML standards require coded mechanisms to interpret information
  - XML is a platform and application neutral data representation language, but leaves document **interpretation up to consumers**
  - XSD and WSDL require **human intervention to appropriately interpret service capabilities** and **information requirements**
  - XSL/T requires pre-built, hand-coded scripts which only enable **syntactic**, **point-to-point** data transformations
- Solutions to these issues have relied on standardized schemas, which do not guarantee cross-organizational interoperability
  - Standardized schemas are difficult to implement
  - Standardized schemas only enforce syntax, not meaning nor usage
  - No single, global schema will meet stakeholder needs across all organizations

Booz | Allen | Hamilton

4

Adoption of organization-specific message formats in a purely Web Services and XML world will impact data interoperability across the ISR COI

- XPath and XSL/T provide point-to-point mappings between a single source and a single target
- Point-to-point mappings between COI-specific message formats will not scale
  - N different formats require N^2 – N mappings
  - Modifications to any single schema require changes to 2N – 2 mappings
  - Tightly-coupled, requiring all involved parties to understand how to interpret everyone else's data
- Tight coupling of XSL/T scripts and mappings violate loose-coupling, a core tenet of Service Oriented Architectures

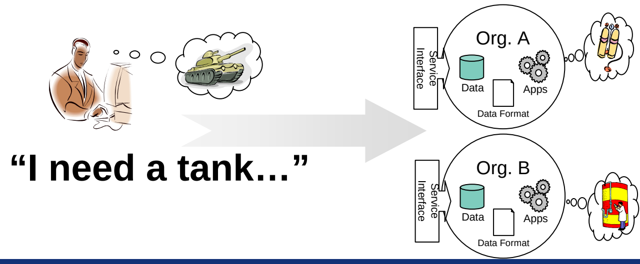


Booz | Allen | Hamilton

5

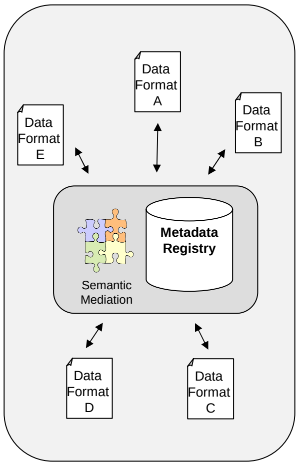
To embrace true data interoperability, mediation infrastructure must provide the ability to interpret and understand data

- Information must become the key foundation for organizations and COIs
  - Data are merely physical values
  - Information is a **meaningful interpretation** of data
- Dynamic information interoperability requires a means interpret the intention and meaning of data
  - Ability to understand the **structure**, **contents**, and **business concepts** embodied in service contracts and message exchanges
  - Ability to **disambiguate** the meaning of similarly named terms



An enhanced mediation infrastructure requires an improved ability for software to interpret message formats

- A loosely-coupled information infrastructure facilitates meaningful interoperability through the use of semantics-based data descriptions
- Semantics-based data descriptions enable a de-emphasis on pre-built, point-to-point mappings
- Mediation infrastructure can transition towards dynamic aggregation and transformation of data by dynamically interpreting data meaning
  - Requires the ability to interpret contents, structure, and meaning of exchanged data
  - Published metadata must describe information contents in an unambiguous, machine-interpretable manner



Achieving Semantic Mediation requires more expressive metadata

- Most forms of metadata focus only on providing syntactic and structural qualities of messages and the services that utilize them

Expressiveness

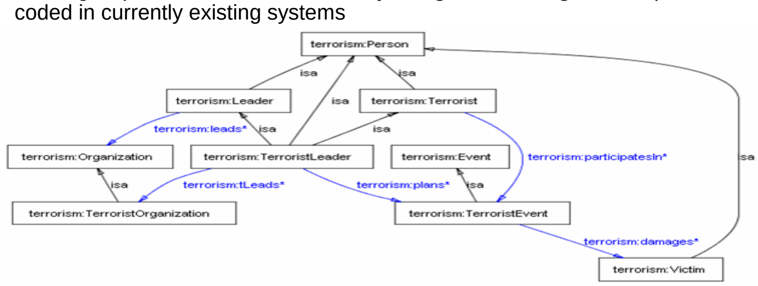
Metadata Type	Description	Examples
Syntactic	Describes the **physical, syntactic markup** of individual data elements (formatting, field markers)	Datatype, Field Length, Field Name, Tag Names, Flat File Makers
Structural	Describes the **logical grouping** of individual of data elements (i.e. entity-attribute groupings)	Logical schema definitions (PersonRecord: PersonName, PersonSSN, PersonDOB)
Semantic	Describes the **codified meaning** of data elements, and their relationships, including any rules or constraints on those relationships	Person **was-born** on PersonDOB, and **was-born** once and only once

- Semantics is the “meaning of data” – the concepts that data represents within a particular context, and the relationships between those concepts.

Booz | Allen | Hamilton

Semantics can be formally modeled in an ontology

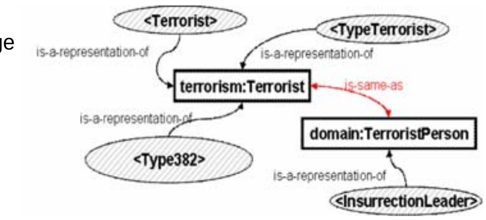
- An ontology is a graph of the abstract concepts, relationships, and logical assertions that comprise a domain
  - Usage and meaning of data are **explicitly** captured in a **machine-interpretable** format
  - Machines can automatically discover **relevant** content sources based on **business concepts**, not just the static labels currently provided by taxonomies
  - Ontologies provide a framework for **exposing** and **reusing** the interpretation **rules** coded in currently existing systems



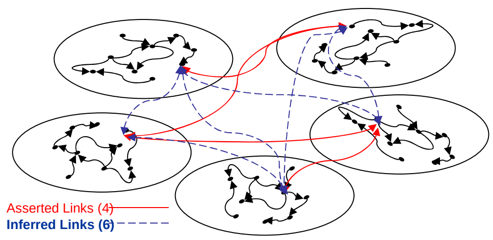
Booz | Allen | Hamilton

Ontologies enable software to meaningfully interpret data, lessening human involvement and increasing efficiency

- Ontologies can be used to bridge other models
  - Relationships can be inferred
  - Schema standardization not required

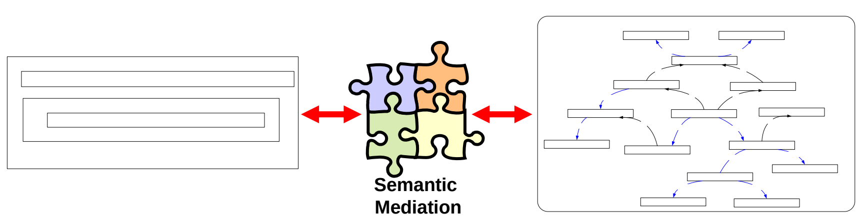


- Ontology constructs can be used to map between ontologies
  - Links are transitive
  - Creates network effect of an enormous scale

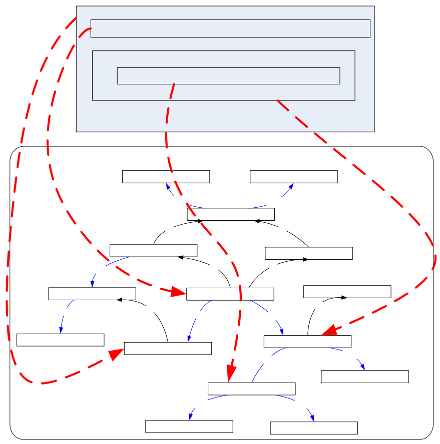


Semantic Data Mediation bridges the gap between the data formats and domain knowledge

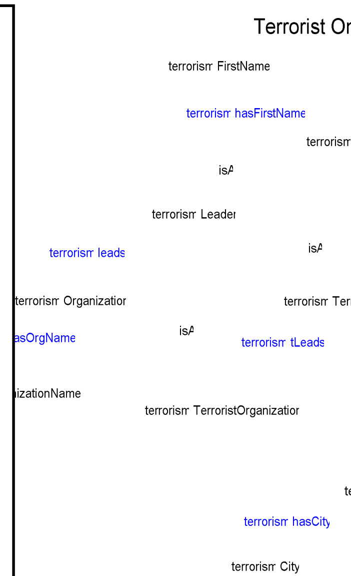
- XML Schema focuses on describing the proper the syntax and structure of a data format
  - Semantic information is implied, but not explicitly codified
  - OWL provides a rich model to define the semantics of a business domain
- Semantic Data Mediation provides a means to autonomously perform dynamic mediations
  - Semantic mappings provide explicit semantic descriptions of data specifications: Concept Entities, Concept Attributes, and Entity Bridges
  - Two-phased approach allows source XML to be recast in OWL for transformation reasoning and exported into target XML



A Concept Entity is a complex-typed XML element that represents a domain business concept

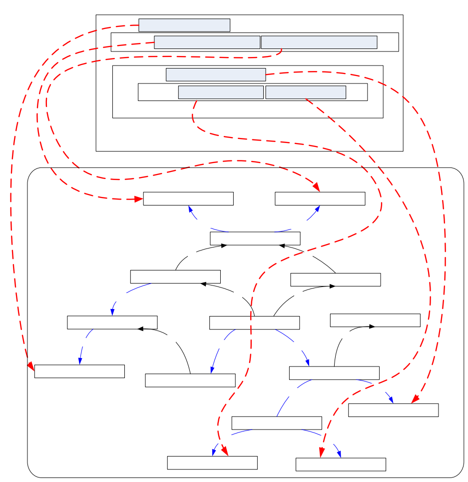


<terrorists>
<group name="Al-Qaida'>
<leader firstName="Osama
<attacks>
<attack date="09/11/200
<target city="New York
</attack>
</attacks>
</group>
</terrorists>



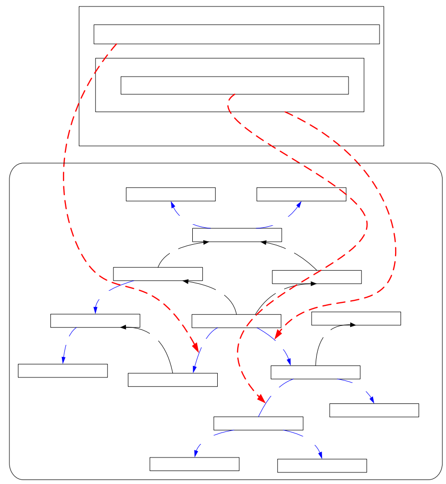
A Concept Attribute is an XML attribute or element that represents a business domain concept, but has a physical value

- Explicitly linked as members of Concept Entities through higher level domain relationships such as **`hasName(Organization, Name)`** or **`hasCity(Location, City)`**



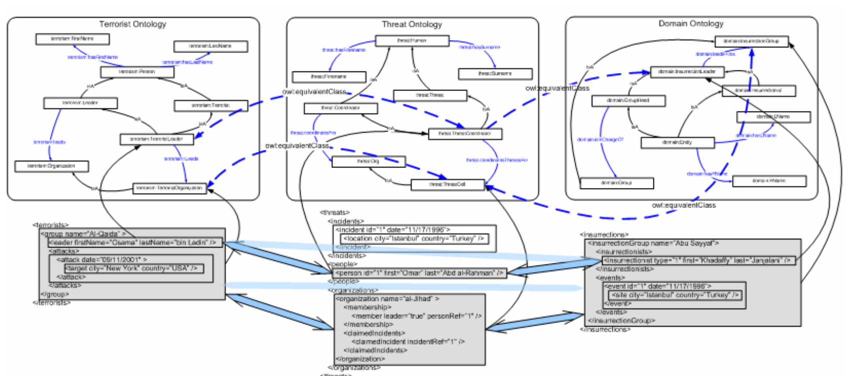
An Entity Bridge represents the higher level domain relationship between two Concept Entities

- Describes how to syntactically and structurally navigate between one the XML element represented by one Concept Entity to another



Inferencing capabilities allow mediation to occur across data specifications that are not directly mapped

- Transitive nature of ontologies provides implicit bridges between semantic data maps
- Reasoning infrastructure able to infer transformation instruction sets

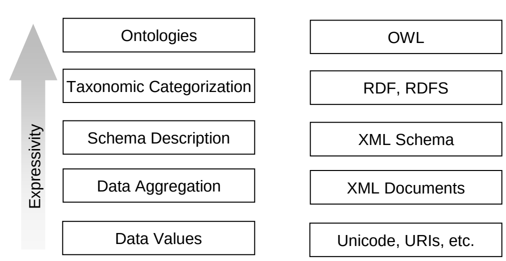


Semantic Mediation techniques codify implicit knowledge to produce explicit information descriptions

Activity	XSLT Mapping/Mediation	Semantic Mapping/Mediation
Concept Extraction	**Implicitly** identify the concept entities, concept attributes, and entity bridges in the source and target XSDs	**Explicitly** document the concept entities, concept attributes, and entity bridges in the source and target XSDs
Structural Navigation	**Implicitly** identify the XPath/XQuery to navigate between the concept entities, concept attributes, and entity bridges for both the source and target XSDs	**Explicitly** document the XPath/XQuery to navigate between the concept entities, concept attributes, and entity bridges for both the source and target XSDs
Semantic Matching	**Manually** identify the semantic likeness of concept entitihes, concept attributes, and entity bridges in the source and target XSDs	Leverage OWL DL reasoning to **autonomously** determine semantic matching between concept entities, concept attributes, and entity bridges
Mediation process	For each source to target XSD mediation, **manually at design-time** compose a stylesheet encompassing the above information	**Dynamically at run-time** generate a semantic mediation between a source and target XSD

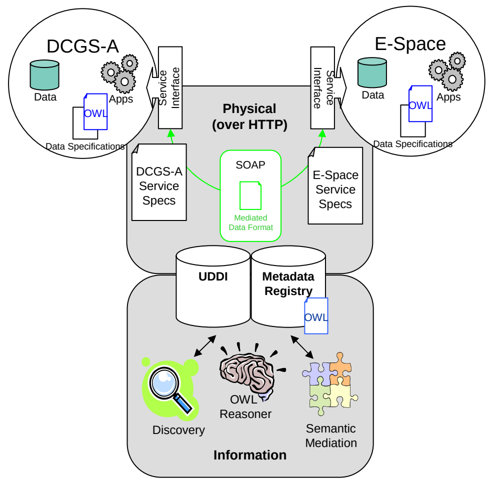
The Semantic Web is a standardized approach towards ontology representation and reasoning that can realize the requirements of Semantic Mediation Infrastructure

- The Semantic Web Activity is a W3C initiative producing standardized mechanisms to specify formal semantics
  - **Resource Description Framework (RDF)**
  - **RDF Schema (RDFS)**
  - **Web Ontology Language (OWL)**
- The Semantic Web stack builds over standard XML and web technologies, easing integration with existing standards

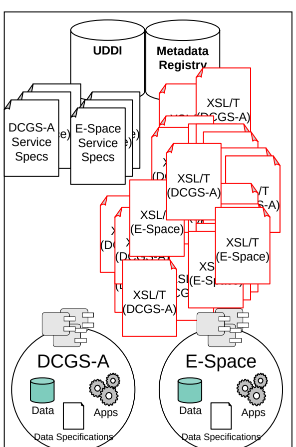


Compatibility with existing web technologies allows Semantic Web technologies to be integrated into a Service Oriented Architecture implementation

- ISR service families, in addition to building localized message formats, can build ISR Domain Ontologies encoded in OWL
- ISR service families describe their XML Schemas using their in OWL
- Cross-ontology mappings leverage existing mappings to relieve any N^2 problems
- ISR organizations register ontologies and OWL-encoded semantic message descriptions
- Semantic Mediation Service interprets registered ontologies and mappings, performing dynamic mediation and fusion

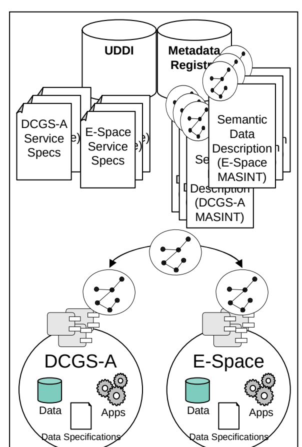


With this computable metadata layer, fewer artifacts are required to support information interoperability in the ISR COI



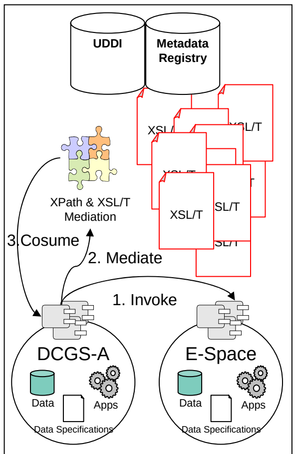
Traditional Web Services SOA

- Traditional Web Services approach:
  - Organization-proprietary specifications for HUMINT, SIGINT, MASINT data
  - Stylesheet mappings required for each permutation of specification integration and fusion
  - Requires up to 30 mappings
- Semantics-enhanced approach:
  - Create domain ontologies describing ISR domains
  - Single ontology bridge between DCGS-A and E-Space
  - 6 total semantic descriptions, one for each message formats



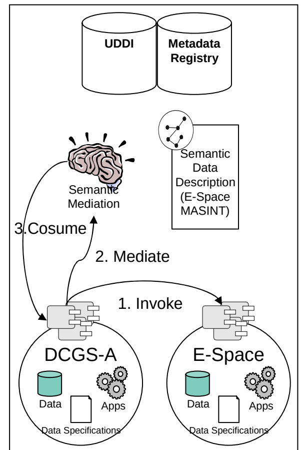
Semantics-enhanced SOA

The semantics-enhanced SOA approach provides a more flexible, scalable mechanism to mediate and consume information



Traditional Web Services SOA

- Traditional Web Services approach:
  - XSL/T Mediation Service resolves a point-to-point mapping
  - Aggregating from multiple sources requires transforming intermediate results
  - Any format change requires 10 mapping modifications
- Semantics-enhanced approach:
  - Semantic Mediation Service resolves a dynamic mediation routine
  - Inferencing over relevant ontologies supports aggregation
  - Any message format change requires 1 mapping modification

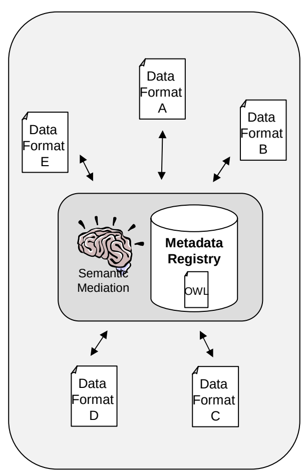


Semantics-enhanced SOA

Booz | Allen | Hamilton

Semantic Mediation can address a proliferation of ISR-related data specifications in an efficient, loosely-coupled manner

- Provide OWL-backed ontological descriptions for data source schemas and content
- Provide ability to enable dynamic, loosely-coupled any-to-any data transformations and aggregations using a semantics-based mediation techniques
- Complexity grows linearly with the number of different data formats
  - Transitive nature of OWL produces a network effect
- Allows organizations to use data formats tailored for their needs, while seamlessly allowing that same data to be shared with the rest of the community



Booz | Allen | Hamilton

Questions?

- For additional information or reference materials, please contact:
  - Sri Gopalan, gopalan_sri@bah.com
  - Sandeep Maripuri, maripuri_sandeep@bah.com
  - Brad Medairy, medairy_brad@bah.com

Booz | Allen | Hamilton

22

Reference Materials

Booz | Allen | Hamilton

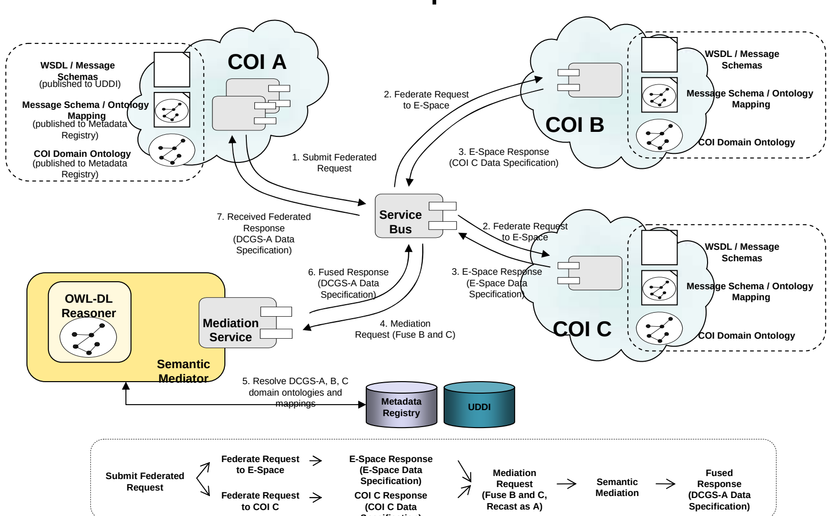
23

Ontologies build beyond taxonomy capabilities by providing a codified, machine-interpretable description of a domain

	Taxonomy	Ontology
Domain Description	Domain categorization based purely on **keywords**	Domain descriptions built through inter-connected **network of relationships** between **domain concepts**
Codified Relationships	Relationships must be **assumed**: offers no mechanism for describing relationships, sub-type or composition:	Relationships are **explicit**: relationship types between concepts are named, and can be **related to other relationships**
Understandability	The significance of each category name must be understood by the consumer to be meaningful	Offers the relationship types to indicate that differently named terms are **equivalent**, **disjoint**, etc.
Intended Consumer	Meant as an organizational system for **humans** to discover and interpret information	Meant as a metadata description framework for **machines** to interpret information
Machine Interpretability	Software must be specifically coded against taxonomy category keywords in order to interpret them	Provides rules to interpret relationships and infer new relationships

Booz | Allen | Hamilton

A semantics-enhanced SOA provides more effective components to realize a traditional Web Services process flow



Booz | Allen | Hamilton

Glossary

- **Service-Oriented Architecture (SOA)** - An application architecture approach in which all functions, or services, are defined using a description language and have invocable interfaces that are called to perform business processes.
- **Web Services -** A standardized way of integrating applications using open standards, such as XML, SOAP, WSDL, and UDDI, over an Internet protocol backbone.
- **SOAP** - A lightweight XML based messaging protocol used to encode the information in web service request and response messages before sending them over a network.
- **Web Services Description Language (WSDL)** – An XML formatted language used to describe a web service's capabilities as collections of communication endpoints capable of exchanging messages.
- **Universal Description, Discovery, and Integration (UDDI)** – A web-based distributed directory that enables businesses to list their services on the internet and discover each other, similar to a traditional phone book's yellow and white pages.

Booz | Allen | Hamilton

Glossary

- **Semantics –** The business meaning and usage of data and services (http://en.wikipedia.org/wiki/Semantics).
- **Ontology –** A domain model specifying real-world concepts and their interrelationships. An ontology is typically characterized by non-attributed entities organized not only by subtyping, hierarchical relationships ('Employee' is-a 'Person'), but additionally by semantic relationships describing how one concept is related to another ('Employee' works-for 'Employer'). Ontologies are commonly used in knowledge representation and artificial intelligence, and are typically used for reasoning, inferencing, and classification computations (http://en.wikipedia.org/wiki/Ontology_%28computer_science%29).
- **Semantic Web –** A W3C project creating a standardized mechanism to enable information exchange by giving meaning, in a manner understandable by machines, to the content of documents on the Web. Semantic Web technologies are not limited to Web-centric hyptertext media, and can be additionally used to describe the meaning and usage of data and services (http://w3c.org/2001/sw).

Booz | Allen | Hamilton

Glossary

- **Resource Definition Framework (RDF) –** An XML-based data model expressing assertions that relate resources (pieces of data) in subject-predicate-object form (RDF Triple). The subject is the ‘thing’ being described, the predicate is the ‘characteristic’ describing the ‘thing’, and the object is the ‘value’ of the ‘characteristic’. This encoding allows software to comprehend sentence-like data assertions (http://www.w3.org/RDF).

- **RDF Schema (RDF/S) –** An RDF-based schema vocabulary language for formally describing groups, or types (known as classes), of RDF resources, and their interrelationships

- **Web Ontology Language (OWL) –** An RDF/S-based ontology language, whose constructs are heavily derived from the DAML+OIL Ontology Language. Adds additional language constructs to provide stronger meaning to RDF/S relationships

- **Reasoning Engine –** A piece of software that attempts to derive answers from a knowledge base. In semantics-based computing, an inference engine typically resolves or discovers interrelationships between ontology classes, allowing conclusions to be drawn about how concepts are related from an underlying ontology.

Booz | Allen | Hamilton